The Sputum Microbiome, Airway Inflammation and Mortality in Chronic Obstructive Pulmonary Disease

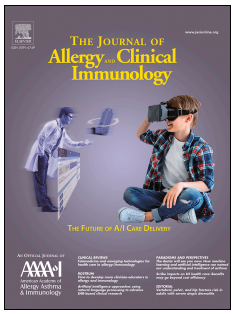
Dicker, Alison J.; Huang, Jeffrey T. J.; Lonergan, Mike; Keir, Holly R.; Fong, Christopher J.; Tan, Brandon

Published in:
Journal of Allergy and Clinical Immunology

DOI:
10.1016/j.jaci.2020.02.040

Publication date:
2021

Journal Pre-proof



The Sputum Microbiome, Airway Inflammation and Mortality in Chronic Obstructive Pulmonary Disease

Alison J. Dicker, PhD, Jeffrey TJ. Huang, PhD, Mike Lonergan, PhD, Holly R. Keir, BSc, Christopher J. Fong, MBChB, Brandon Tan, MBChB, Andrew J. Cassidy, PhD, Simon Finch, MBChB, Hana Mullerova, PhD, Bruce E. Miller, PhD, Ruth Tal-Singer, PhD, James D. Chalmers, PhD

PII: S0091-6749(20)30566-2

DOI: https://doi.org/10.1016/j.jaci.2020.02.040

Reference: YMAI 14519

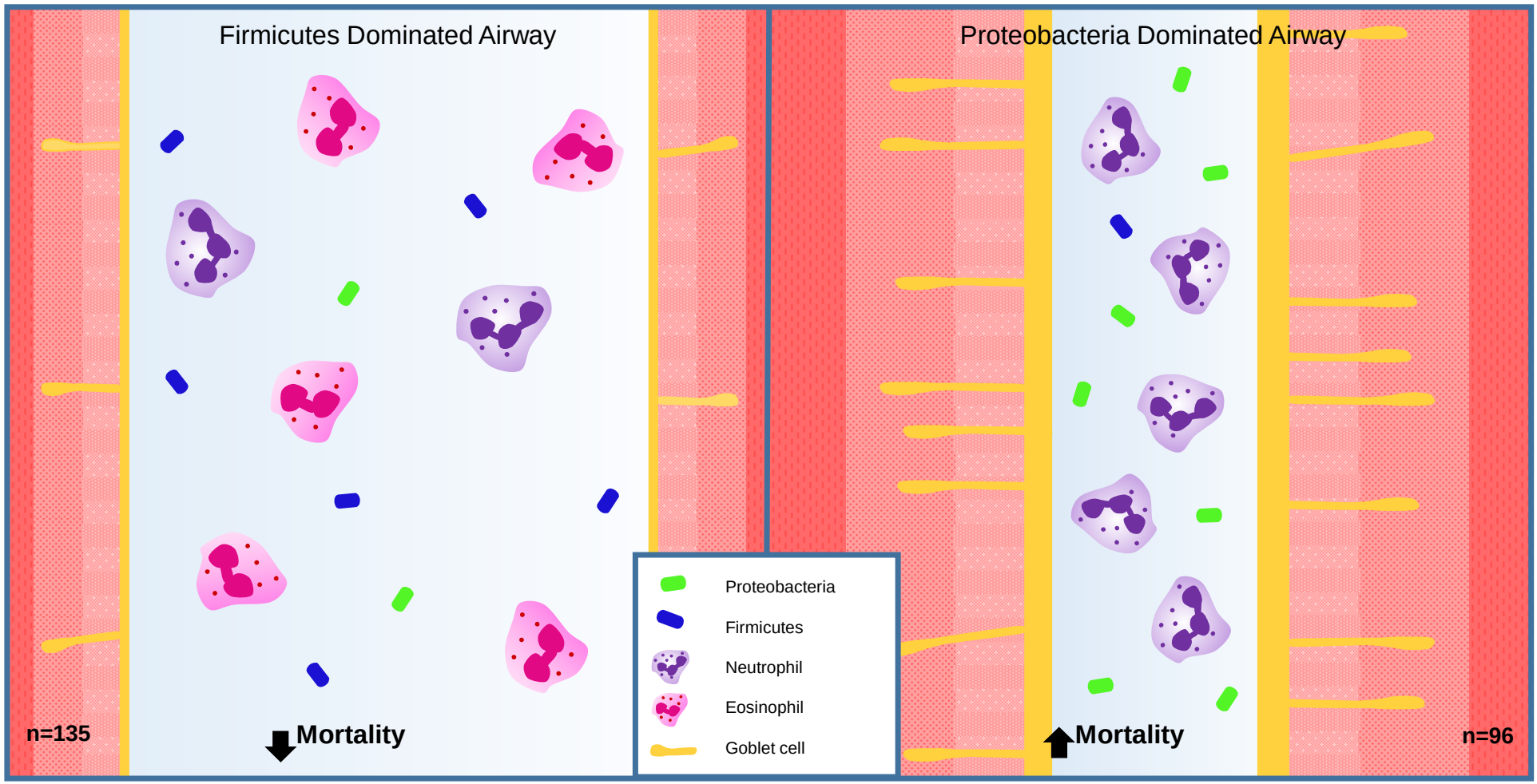
To appear in: *Journal of Allergy and Clinical Immunology*

Received Date: 26 October 2019

Revised Date: 17 February 2020

Accepted Date: 18 February 2020

Please cite this article as: Dicker AJ, Huang JT, Lonergan M, Keir HR, Fong CJ, Tan B, Cassidy AJ, Finch S, Mullerova H, Miller BE, Tal-Singer R, Chalmers JD, The Sputum Microbiome, Airway Inflammation and Mortality in Chronic Obstructive Pulmonary Disease, *Journal of Allergy and Clinical Immunology* (2020), doi: https://doi.org/10.1016/j.jaci.2020.02.040.


Microbiome, Airway Inflammation and Mortality in COPD
n=253
Firmicutes Dominated Airway
Proteobacteria Dominated Airway
Proteobacteria
Firmicutes
Neutrophil
Eosinophil
Goblet cell
n=135
Mortality
Mortality
n=96

The Sputum Microbiome, Airway Inflammation and Mortality in Chronic Obstructive Pulmonary Disease

Alison J Dicker PhD[1], Jeffrey TJ Huang PhD[1], Mike Lonergan PhD[1], Holly R Keir BSc[1], Christopher J Fong MBChB [1], Brandon Tan MBChB[1], Andrew J Cassidy PhD[1], Simon Finch MBChB[1], Hana Mullerova PhD[2], Bruce E Miller PhD[2], Ruth Tal-Singer PhD[2], James D Chalmers PhD[1]

Scottish Centre for Respiratory Research, University of Dundee, United Kingdom[1];

Medical Innovation, GSK R&D, Collegeville, PA, USA[2]

Corresponding author: James D Chalmers, Scottish Centre for Respiratory Research, University of Dundee,

Ninewells Hospital and Medical School, Dundee, DD1 9SY, jchalmers@dundee.ac.uk, +44 1382 386131

Conflicts of interest

JDC reports grants from Glaxosmithkline (GSK) for this research. In addition, he reports grants and personal fees from Glaxosmithkline, Boehringer-Ingelheim, AstraZeneca, Pfizer, Bayer Healthcare, Grifols, Napp, Aradigm corporation, and Insmed outside the submitted work. HM and RTS were employees of GSK at the time this study was conducted. RTS is a shareholder of GSK. BEM is an employee and shareholder of GSK. The other authors declared no conflicts of interest.

Word count: 3311

Abstract

Background: The sputum microbiome has a potential role in disease phenotyping and risk stratification in chronic obstructive pulmonary disease but few large longitudinal cohort studies exist.

Objective: To investigate the COPD sputum microbiome and its association with inflammatory phenotypes and mortality.

Methods: 16S rRNA gene sequencing was performed on sputum from 253 clinically stable COPD patients (4-years median follow-up). Samples were classified as Proteobacteria or Firmicutes (phylum level) and *Haemophilus* or *Streptococcus* (genera level) dominant. Alpha diversity was measured using Shannon-Wiener Diversity and Berger-Parker Dominance Indices. Survival was modelled using Cox proportional hazards regression. A subset of 78 patients had label-free liquid chromatography/mass spectrometry performed, with partial least square discriminant analysis integrating clinical, microbiome and proteomics data.

Results: Proteobacteria dominance and lower diversity was associated with more severe COPD using the GOLD classification system (p=0·0015), more frequent exacerbations (p=0·0042), blood eosinophils ≤100cells/μL (p<0·0001) and lower FEV_1 (p=0·026). Blood eosinophil counts showed a positive relationship with %Firmicutes and *Streptococcus*, and a negative association with %Proteobacteria and *Haemophilus*. Proteobacteria dominance was associated with increased mortality compared to Firmicutes dominated or balanced microbiome profiles (HR 2·58 95%CI 1·43-4·66, p=0·0017 and HR 7·47 95%CI 1·02-54·86, p=0·048 respectively). Integrated omics analysis showed significant associations between Proteobacteria dominance and the neutrophil activation pathway in sputum.

Conclusion: The sputum microbiome is associated with clinical and inflammatory phenotypes in COPD. Reduced microbiome diversity, associated with Proteobacteria (predominantly *Haemophilus*) dominance, is associated with neutrophil associated protein profiles and an increased risk of mortality.

Capsule summary: Microbiome analysis of sputum reveals distinct endotypes of COPD with different inflammatory profiles and long term survival.

Key Messages:

- The sputum microbiome in COPD is associated with clinical and inflammatory phenotypes

- Dominance and loss of microbial diversity are associated with increased mortality in COPD
- Microbiome and inflammatory profiles are linked, with Proteobacteria dominant profiles being associated with neutrophil activation markers, and Firmicutes dominant profiles being associated with raised blood eosinophil counts.

Keywords: Microbiome, COPD, Eosinophil, Phenotype, *Haemophilus,*

Abbreviations

aHR:	Adjusted hazards ratio
BPDI:	Berger-Parker dominance index
BMI:	Body mass index
CAT:	COPD assessment test
COPD:	Chronic Obstructive Pulmonary Disease
DNA:	Deoxyribonucleic acid
FDR:	False discovery rate
FEV_1:	Forced expiratory volume in 1 second
FVC:	Forced vital capacity
GOLD:	Global initiative for chronic obstructive lung disease
HR:	Hazard ratio
ICS:	Inhaled corticosteroids
LC-MS/MS:	Liquid chromatography with tandem mass spectrometry
MRC:	Medical research council
OTU:	Operational taxonomic unit
QIIME:	Quantitative insights into microbial ecology
RNA:	Ribonucleic acid
SGRQ:	St Georges respiratory questionnaire
SWDI:	Shannon-Wiener diversity index
TARDIS:	Tayside allergic and respiratory disease information system

Introduction

Chronic Obstructive Pulmonary Disease (COPD) is recognized as a heterogeneous disease with patients manifesting multiple diverse phenotypes, endotypes and "treatable traits".(1-4) Recent therapeutic concepts in COPD have moved away from the previous "one size fits all" treatment approach with bronchodilators, inhaled corticosteroids (ICS) and other medications, towards a precision medicine approach in which clinical characteristics and biomarkers are used to direct treatments and to optimize their risk: benefit profile.(5-8) A recent example of this is the use of blood eosinophil counts to direct treatment with inhaled corticosteroids.(7, 8) The use of a blood biomarker to guide treatment has gained acceptance due to identification of a relationship between blood and sputum eosinophil counts. Randomized controlled trials have subsequently shown that patients with blood eosinophil counts less than 100cells/µL derive no significant benefit in terms of exacerbation reduction when treated with ICS, and may be at increased risk of pneumonia.(6, 9, 10)

Eosinophilic COPD is an example of an endotype, a subtype of a condition defined by a distinct pathophysiological mechanism.(2) Inflammation, however, is only one of the pathophysiological features of COPD which is associated with airway structural damage and airway infection.(11, 12) Bacteria are thought to play a key role in COPD pathogenesis and recent data has brought a deeper understanding of the complex microbial communities present in the lung affected by COPD.(11, 13-16) Studies of the COPD "lung microbiome" to date have found that as COPD becomes more severe in terms of lower lung function, there is an increase in the relative abundance of Proteobacteria, and particularly the genus *Haemophilus*.(11, 13, 15) Lower microbiome diversity is also associated with increased severity of disease and neutrophilic inflammation including the formation of neutrophil extracellular traps.(11, 13, 15)

Exacerbations are key events in the natural history of COPD and recent work suggests no consistent pattern of altered microbial profiles at exacerbation compared to stable disease.(15) Previous work suggests there are at least four different exacerbation "endotypes" associated with distinct inflammatory profiles - bacterial exacerbations, eosinophilic exacerbations, virus predominant exacerbations and pauci-inflammatory events.(17) These events, defined by clinical and inflammatory criteria, may also have distinct microbiome profiles.(18)

Lower lung function and frequent exacerbations may relate to endotypes linked to greater mortality in COPD.(19-21) If changes in the microbiome loss of microbial diversity are indeed related to more severe disease and high risk patient subtypes, it is plausible that the respiratory microbiome may associate with

long term mortality as has been observed in idiopathic pulmonary fibrosis.(22) No studies have assessed the association of the sputum microbiome profile with mortality in stable COPD.

Here we report the results from a longitudinal cohort study of COPD patients designed to integrate data from the sputum microbiome and proteome with clinical phenotypes and long-term clinical outcomes, to provide further insights into the relevance of the respiratory microbiome in clinical practice.

Methods

Study Design

We performed a longitudinal observational study of patients with a diagnosis of COPD nested within a population based COPD registry (Tayside Allergic and Respiratory Disease Information System (TARDIS)) in the East of Scotland.(16, 20) Patients were invited to participate in a microbiome sub-study and were included if they were >40 years, had a FEV_1/FVC ratio <70% at screening, and a clinical diagnosis of COPD. Exclusion criteria included the inability to give informed consent; primary diagnosis of asthma; and systemic immunosuppression (excluding prednisolone at 5mg or less daily). Additionally, patients needed to be clinically stable and free of antibiotic or oral corticosteroid therapy for > 4 weeks prior to enrolment. All relevant medical history (comorbidities, current medications, significant past conditions, operations and diagnostic procedures) was recorded at screening. Participants provided induced sputum samples following induction with 3% hypertonic saline. Spirometry, St Georges Respiratory Questionnaire (SGRQ), COPD assessment test (CAT) and MRC dyspnoea scoring was performed at each visit. Exacerbations in the study period were defined as an increase in respiratory symptoms greater than day to day variation requiring a change in therapy; participants returned to the clinic for assessment and were given a standardized treatment of antibiotics and corticosteroids. Patients were enrolled from 2013-2015 and survival data was obtained from linked medical records within the TARDIS registry as described previously.(16)

Clinical phenotypes

A-priori, we identified clinical phenotypes that have been linked with response to therapies in COPD. These were low blood eosinophil counts (examined using a cut-off of ≤100cells/μL according to recent Global Initiative for Chronic Obstructive Lung Disease (GOLD) recommendations and linked to ICS response),(23) chronic bronchitis (defined as daily sputum production at least three months of the year over at least two years and linked to response to treatments including roflumilast) and, finally, the frequent exacerbator phenotype, defined as patients with ≥ 2 exacerbations/year.(9, 24) Clinical phenotypes were not mutually exclusive.

Sputum Microbiome

DNA was extracted from whole sputum as described in the online supplement, followed by 16S rRNA gene sequencing on the Illumina MiSeq platform. Bioinformatic analysis and quality checking of the resulting sequences was performed using QIIME (version 1.9.0) and R (version 3.4.0). Shannon-Wiener Diversity Index (SWDI, where a higher value indicates a sample is more diverse) and Berger-Parker

Dominance Index (BPDI, which measures the proportion of the microbiome dominated by the most abundant taxa, with a value closer to 1 indicating a microbiome with greater dominance) were used as measures of alpha diversity of samples. See online supplement for full methods.

Characterization of microbiome subtypes of COPD

Prior studies of the COPD sputum microbiome identified that Proteobacteria and Firmicutes were the dominant phyla and that inflammation and COPD severity were strongly associated with the presence of a single genera at >40% relative abundance in many patients.(11, 13, 18) Therefore *a-priori* we defined candidate microbiome subgroups. At the phylum level samples were classified as either Proteobacteria dominant or Firmicutes dominant based on >40% Observed Taxonomic Units (OTUs) of either phylum. For analysis at the genera level we classified samples into *Haemophilus* dominant and *Streptococcus* dominant (as the most abundant Firmicute); "balanced" microbiome profiles were defined by the absence of >40% OTUs of either Proteobacteria or Firmicutes at the phylum level or no individual OTU exceeding 40% at the genera level. The cut-off of 40% OTUs to define these groups was used based on our previous work.(11) Throughout the manuscript we refer to the respiratory microbiome to reflect that all assessments are made on sputum which includes contributions from the lower and upper respiratory tracts.

Integration of microbiome, clinical and proteomic data

A subset of 78 patients included in the primary study provided sufficient sputum samples for proteomics with the results integrated with their microbiome and clinical data. The total protein concentrations of sputum supernatants were quantified using Pierce 660 protein assays. Sputum protein (50µg) from each sample was added to an equal volume of acetonitrile before incubating at 100°C for 15mins. The samples were dried down in a centrifugal vacuum and resuspended with 50 mM ammonium bicarbonate (pH 8·5) to a final concentration of 1mg/mL. Samples were then reduced and alkylated before subjecting to nano-flow-LC-MS/MS analysis. Protein identification and label-free quantification were carried out using Maxquant (version 1.4.1.2) against Uniprot-human database (version 2014-07-09). The fixed modification was carbamidomethylation on cysteine, and variable modifications include oxidation on methionine and N-terminal acetylation. False Discovery Rate (FDR) for protein identification was set to 1% at protein level.

Statistical analysis

Statistical analysis of data was carried out using R version 3.4.0, SPSS version 21 and GraphPad Prism 6.07. Continuous data are presented as median (interquartile range) and categorical data are presented as N (%). Continuous parametric data were compared using unpaired T-test while continuous non-

parametric data were compared using the Mann-Whitney U test. Generalized linear models with binomial errors and a logistic link function were used to model the relationship between blood eosinophils and relative abundance. Principal component analysis was used to analyse microbiome beta diversity. Unadjusted survival was studied using Kaplan-Meier survival analysis. Cox Proportional Hazards models incorporating age, sex, baseline FEV_1 and exacerbation frequency in the previous year were used to model survival with the proportional hazards assumption checked using log-minus-log plots. Integrated omics analysis was performed using partial least square discriminant analysis with FDR correction using the Benjamini-Hochberg method, followed by a separate univariate analysis to confirm the results and to avoid overfitting. Statistical significance was set at $P<0{\cdot}05$ for all other analyses.

Results

296 participants were enrolled in the study; after excluding 44 patients whose baseline sputum samples failed DNA extraction and sequencing quality controls the final cohort was 252 participants (Figure 1) with the baseline characteristics of the cohort shown in Table 1. The predominant phyla observed in the microbiome were Proteobacteria and Firmicutes, with fewer OTUs identified as *Bacteroidetes*, *Actinobacteria* and *Fusobacteria* (Figure 2A). At the genus level the most abundant genera were *Haemophilus*, *Streptococcus*, *Neisseria*, *Veillonella*, and *Prevotella*, whilst some individual samples additionally showed a predominance of genera such as *Pseudomonas*, *Stenotrophomonans*, *Pasteurella*, Unknown *Enterobacteriaceae*, and *Moraxella* (Figure 2B).

The non-eosinophilic phenotype has lower alpha diversity and increased relative abundance of Proteobacteria

We observed a clear relationship between microbiome and clinical phenotypes. Proteobacteria dominated microbiomes (n=96) were associated with poorer lung function and more frequent exacerbations. Significantly more *Streptoccocus* dysbiotic microbiomes (n=56) were GOLD group D (p=0·011) and had a worse mMRC score compared to balanced genera (p=0·0017) (Figure S2).

The low blood eosinophil subgroup (n=62) had lower microbiome diversity as measured by SWDI (p<0·0001) whilst the BPDI was higher in the low blood eosinophil subgroup (p<0·0001). These differences appeared to be associated with a higher relative abundance of Proteobacteria at the phylum level (p=0·0001) and *Haemophilus* at the genus level (p=0·046, Figure 3B). We observed a statistically significant negative correlation between % Proteobacteria and eosinophil count in peripheral blood (Spearman r correlation, p<0·0001, Figure 3C) and a positive correlation between % Firmicutes and eosinophil count (p<0·0001, Figure 3C). Further data are shown in Figure S3 online.

Differences at both the phylum and genus level were observed when comparing patients who frequently exacerbate compared to non-frequent exacerbators (defined as those with ≤2 exacerbations/year) and comparing patients with chronic bronchitis compared to those patients without chronic bronchitis (Figure S4). Principle component analysis suggested that there was no clear clustering of microbiome based on either blood eosinophil group, chronic bronchitic status or exacerbation frequency.

Microbiome profiles are linked to long term mortality

50 participants (19·8% of patients enrolled) died during a median follow-up of four years. At the genera level, *Haemophilus* dominated profiles (HR 2·53 95%CI 1·08-5·94, p=0·032) were associated with increased mortality compared to balanced profiles. *Streptococcus* dominated profiles were associated with

a similar mortality risk to *Haemophilus* when compared to balanced profiles (HR 2·05 95%CI 0·80-5·20, p=0·13 Figure 4). The adjusted hazard ratios (aHR) were similar to the primary analysis (aHR 2·17 95%CI 0·89-5·27, p=0·09 and aHR 1·67 95%CI 0·64-4·37, p=0·30 for *Haemophilus* and *Streptococcus* dominated profiles respectively).

At the phylum level, Proteobacteria dominance was associated with increased mortality compared to both Firmicutes dominated and balanced microbiome profiles (Hazard ratios (HR) 2·58 95%CI 1·43-4·67, p=0·0017 and 7·47 95%CI 1·02-54·86, p=0·048, respectively). The relationship between Proteobacteria dominance and survival compared to Firmicutes persisted after multivariate adjustment for age, sex, FEV_1 and previous exacerbation frequency, aHR 2·35 95%CI 1·27-4·32, p=0·006. Firmicutes dominated and balanced profiles were more similar to each other (HR 2·85 95%CI 0·38-21·43, p=0·31) (Figure 4). As *Streptococcus* dominated profiles appeared to behave differently to others in the Firmicutes phylum, a sensitivity analysis was conducted comparing Proteobacteria dominated profiles to Firmicutes dominated profiles with *Streptococcus* excluded (Figure S5). This did not greatly affect the results, with a clearly worse mortality associated with Proteobacteria dominance (HR 2·93 95%CI 1.22-7·06 p=0.017, aHR 2·80 95%CI 1·13-6·90 p=0·026.

Lower alpha diversity was also strongly associated with long term outcome. The median SWDI for all samples was 3·36. Patients with values below this level had a markedly increased mortality (HR 4·05 95%CI 2·07-7·93, p<0·0001) which persisted after multivariate adjustment (aHR 3·08 95%CI 1·57-6·09, p=0·001). The median BPDI was 0·41 and values above this level were similarly predictive of mortality (HR 3·81 95%CI 1·99-7·30, p<0·0001), aHR 3·36 95%CI 1·73-6·51, p=0·0003) (see online Figure S6).

Integration of clinical, microbiome and proteomic data shows that Proteobacteria dominated microbiome profiles are associated with neutrophilic inflammation

To investigate how microbiome profiles influence pathophysiology of COPD, we further investigated the sputum protein profiles by nano-flow-LC-MS/MS from a subset of patients (n=78) with three predefined (Proteobacteria dominated (33), balanced (15), Firmicutes dominated (30)) sputum microbiome types. All patients with Proteobacteria and Firmicutes dominated profiles had greater than 50% OTUs of these phyla in their sputum sample. The characteristics of these patients are shown in Supplementary Table S1.

The combined dataset consisted of 113 taxa identified through the microbiome analysis, 21 clinical variables and 613 proteins. Partial least square discriminant analysis of the combined dataset revealed separation between Proteobacteria, balanced and Firmicutes microbiome profiles (R^2X=0·17, R^2Y=0·73, Figure 5A). The loadings plot (Figure 5B, used to determine what variable(s) in the dataset

drive the separation between the different microbiome profiles) indicated that the Proteobacteria dominated microbiome cluster was associated with multiple proteins including myeloperoxidase, catalase, matrix metalloproteinase 9 and 8, and neutrophil elastase, all of which may be associated with neutrophilic inflammation. Further pathway analysis indicated that significantly upregulated proteins (Supplementary Table S2) associated with the Proteobacteria group are over-represented within the "neutrophil activation" pathway (p=2·2E-14, FDR corrected) adding further evidence to a possible association between Proteobacteria and neutrophilic inflammation. In contrast, the Firmicutes dominated microbiome cluster was associated with proteins such as Cystatin B (CSTB), Folate Receptor 1(FOLR1), Small Proline Rich Protein 3(SPRR3), Golgi Membrane Protein 1 (GOLM1), and Clusterin (CLU) (Figure 5B, black labels). Pathway analysis of significantly upregulated proteins (Supplementary Table S2) in the Firmicutes group showed an over-representation of the "negative regulation of peptidase activity" pathway (p=3·7E-4, FDR corrected) including cystatin B, Cysteine-S, alpha 1 antitrypsin, serpin B3 and WAP four disulphide core domain protein 2.

Discussion

To the authors' knowledge, this is the first study to demonstrate that the sputum microbiome in stable COPD patients is associated with long term survival in a prospective cohort of patients; a recent study has linked the microbiome (reduced alpha diversity, increased *Staphylococcus* and reduced *Veillonella*) at acute exacerbation of COPD requiring hospitalization with increased mortality.(25) Taken together these studies suggest that the respiratory microbiome is associated with long term outcomes in COPD. In our study, sputum microbiome profiles at baseline were linked with clinical phenotypes, with exacerbation subtypes and ultimately with long term outcomes. Proteobacteria dysbiosis, defined by the dominance of one or more organisms including well recognized COPD pathogenic genera such as *Haemophilus*, *Moraxella* and *Pseudomonas*, was more frequent in patients with low blood eosinophil counts, chronic bronchitis symptoms and patients with frequent exacerbations.(26-28) These patients had worse lung function and ultimately increased mortality. In contrast, patients with Firmicutes dominance had milder disease, apart from those patients with dominance of the genera *Streptococcus.* Patients with dominance due to *Streptococcus* had a high level of disease severity with impairment of quality of life, lung function impairment and had equivalent mortality to patients with *Haemophilus* dominance. Neither *Haemophilus* nor *Streptococcus* was significantly associated with mortality compared to balanced profiles after adjustment for age, sex, FEV_1 and exacerbation history. In the case of *Haemophilus*, the HR was still greater than 2 after multivariable adjustment suggesting our study may have been underpowered. IWe, and others, have previously shown an association between Proteobacteria, *Haemophilus* and neutrophilic inflammation.(11, 29) We have now extended this observation by conducting an integrated analysis of microbiome, proteomic and clinical data which demonstrates a clear statistically significant association between Proteobacteria and neutrophilic inflammatory markers such as neutrophil elastase, myeloperoxidase and matrix metalloproteinases which are released from neutrophils during inflammatory responses. Although the pathway analysis and prior work suggests these are neutrophil associated, other cells such as macrophages may contribute to the release of matrix metalloproteinases and myeloperoxidase for example. Our data therefore suggests that while there is significant overlap in the phenotypic characteristics of patients with COPD, at each extreme of the spectrum are patients with Proteobacteria dominant profiles with eosinopenia, chronic bronchitis, neutrophil dominated inflammation, frequent exacerbations, poor lung function and reduced survival, and conversely *Streptococcus* dominant profiles associated with raised blood eosinophil counts, the absence of chronic bronchitis and frequent exacerbations. Whether these microbiome profiles link to response to treatments such as inhaled corticosteroids or anti-inflammatory drugs remains to be established but neutrophilic disease is established to be less responsive to inhaled corticosteroids.(12)

It is not possible from our analysis to disentangle cause and effect when considering the impact of the microbiome on clinical phenotype and clinical outcomes. Proteobacteria dysbiosis may be associated with a more rapid decline in FEV_1, or changes in the lung during remodelling may predispose to Proteobacteria dominance, or both statements may be true. Similarly, patients with raised eosinophil counts have an excess of *Streptococcus*. This may suggest that eosinophilic inflammation predisposes to *Streptococcus* infection/ colonisation, as has been demonstrated for other Firmicutes such as *Staphylococcus*,(30) or some members of the *Streptococcus* genera may provoke an eosinophilic response, or changes in the microbiome may be the result of external factors such as antibiotic treatment, inhaled or oral corticosteroid therapy or interactions with the upper airway and gut microbiomes.(31) Further mechanistic studies will be required to test these above hypotheses and to determine the cause or effect of the association between eosinophil count and *Streptococcus*. The cross-sectional nature of our study makes causal inference impossible. Antibiotic treatment in particular is an important potential confounder. The relationship between diversity and the frequent exacerbator phenotype may reflect the effect of repeated antibiotic courses in patients with frequent exacerbations. Some aspects of our survival analysis nevertheless suggest that dominance and loss of microbial diversity themselves may be harmful and contribute to disease progression. The strength of association with survival was striking with a more than 2-fold increased risk of death among individuals with Proteobacteria or *Haemophilus* dominance. While this could reflect these patients simply having more severe disease, adjustment for age, sex, and FEV_1 and exacerbation frequency, did not significantly modify the association of Proteobacteria with mortality. We have previously demonstrated that *Haemophilus* dysbiosis is associated with an increased frequency of exacerbations and increased airway neutrophil extracellular trap formation.(11) NET formation exposes the airway to increased concentrations of toxic proteases and antimicrobial peptides such as neutrophil elastase and matrix metalloproteinases which are linked to disease progression in emphysema and COPD, and which were linked to Proteobacteria dominance in this study.(32) There is therefore a clear causal pathway through which bacterial infection could lead to disease progression and increased mortality. Despite the strength of our findings the number of events during follow-up was relatively small and confidence intervals wide. Our findings should ideally be replicated in future longitudinal microbiome studies.

A series of microbiome studies of varying designs, including cross-sectional and longitudinal cohorts have now provided remarkably consistent results. The dominance of *Haemophilus* and *Streptococcus* as key taxa in COPD, distinct microbiome subtypes of exacerbation, the association with lower diversity and poor lung function and more severe disease are all highly consistent across multiple cohorts.(11, 13, 15, 18) This requires a consideration of how the microbiome could be therapeutically targeted. Here there is an absence of evidence. *Haemophilus influenzae*, the most frequently identified species on culture in

those with *Haemophilus* dominance by sequencing, may be amenable to treatment with antibiotics but the microbiome has been shown to be remarkably resistant to short term disruption with antibiotics.(11, 15) Long term antibiotic treatment such as with macrolides has been shown to modify the microbiome, but in a study by Rogers *et al*, in patients with bronchiectasis, a reduced relative abundance of *Haemophilus* was associated with an increased relative abundance of other Proteobacteria including *Pseudomonas*.(33) Therefore it is not certain that antibiotic treatment would produce a positive change in the microbiome. Vaccination, pulmonary or gastrointestinal probiotics or discontinuation of immunosuppressive medications such as corticosteroids have all been considered but interventional studies are now required to establish if treatments can modify the microbiome in a meaningful way.

Limitations of our study should be acknowledged. Our study was enrolled from a single UK region and results may not be generalizable to other settings worldwide. Our study included a high proportion of patients with severe COPD and frequent exacerbations and consequently a high frequency of use of preventative therapies such as inhaled corticosteroids and antibiotics. CT scanning can identify additional phenotypic characteristics in COPD such as those patients with dominant bronchiectasis and emphysema and the absence of systematic CT scanning in our study is an important limitation. We reported all-cause mortality but did not have sufficient data to examine associations with respiratory specific mortality. We used 16S rRNA sequencing to characterize the respiratory microbiome. This is the most widely used method in the literature but has inherent biases which must be considered in interpretation of the results. The most relevant limitation in this case is the lack of resolution to determine organism identity to species level. It is therefore not known which Streptococcal species are being identified in the *Streptococcus* cluster; use of species specific PCR could resolve the identity to species level in future studies.(34) It should also be noted that sputum, while widely used in studies of the airway microbiome, has limitations in that it is often intermediate between bronchoalveolar lavage and upper airway swabs, therefore containing contributions from the upper and lower airway. Since our study aims to test the prognostic and phenotypic utility of the sputum microbiome, we make no inferences about lung ecology. Whilst different studies have utilized various cut-offs to investigate associations between microbiome profiles and blood eosinophil counts (e.g. 2%), we used 100cells/µL to define eosinophilia, based on the recent GOLD 2019 clinical guidelines. Our generalised linear models show a continuous relationship between microbiome profiles and blood eosinophil counts and did not suggest a true “cut-off” which is consistent with current thinking within the COPD field as a whole. Our proteomic analysis was only performed in a subset of patients. While the characteristics of these patients were similar to the overall cohort this introduces a risk of selection bias. Our study has many strengths including the large sample size compared to other COPD microbiome studies, the inclusion of a representative cohort including patients across the spectrum of COPD and the availability of long-term microbiome samples and follow-up data.

In conclusion, we have identified microbiome associated subtypes of COPD associated with clinical phenotypes and increased mortality. Our results support a personalized medicine approach to therapy in COPD.

Acknowledgements

We acknowledge Mr Alun Barton at the Biomarker and Drug Analysis Core Laboratory (University of Dundee) for technical assistance and Dr. Robert Dickson for his helpful comments on the manuscript.

Contributorship section

Study design: AJD, HM, BEM, RTS, JDC
Data collection: AJD, SF, JDC
Laboratory analysis: AJD, JTJH, HRK, CJF, BT, AJC, SF
Data analysis: AJD, JTJH, ML, BEM, RTS, JDC
Data interpretation: AJD, JTJH, ML, BM, RTS, JDC
Drafting the manuscript: AJD, JTJH, BM, RTS, JDC
Revising the manuscript and approval for submission: All authors

379

References

1. Agusti A, Bel E, Thomas M, Vogelmeier C, Brusselle G, Holgate S, et al. Treatable traits: toward precision medicine of chronic airway diseases. Eur Respir J. 2016;47(2):410-9.
2. Anderson GP. Endotyping asthma: new insights into key pathogenic mechanisms in a complex, heterogeneous disease. Lancet. 2008;372(9643):1107-19.
3. Han MK, Agusti A, Calverley PM, Celli BR, Criner G, Curtis JL, et al. Chronic obstructive pulmonary disease phenotypes: the future of COPD. Am J Respir Crit Care Med. 2010;182(5):598-604.
4. Agusti A, Bafadhel M, Beasley R, Bel EH, Faner R, Gibson PG, et al. Precision medicine in airway diseases: moving to clinical practice. Eur Respir J. 2017;50(4).
5. Chalmers JD, Tebboth A, Gayle A, Ternouth A, Ramscar N. Determinants of initial inhaled corticosteroid use in patients with GOLD A/B COPD: a retrospective study of UK general practice. NPJ Prim Care Respir Med. 2017;27(1):43.
6. Vogelmeier CF, Criner GJ, Martinez FJ, Anzueto A, Barnes PJ, Bourbeau J, et al. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Lung Disease 2017 Report: GOLD Executive Summary. Eur Respir J. 2017;49(3).
7. Brusselle G, Pavord ID, Landis S, Pascoe S, Lettis S, Morjaria N, et al. Blood eosinophil levels as a biomarker in COPD. Respir Med. 2018;138:21-31.
8. Pascoe S, Locantore N, Dransfield MT, Barnes NC, Pavord ID. Blood eosinophil counts, exacerbations, and response to the addition of inhaled fluticasone furoate to vilanterol in patients with chronic obstructive pulmonary disease: a secondary analysis of data from two parallel randomised controlled trials. Lancet Respir Med. 2015;3(6):435-42.
9. Lipson DA, Barnhart F, Brealey N, Brooks J, Criner GJ, Day NC, et al. Once-Daily Single-Inhaler Triple versus Dual Therapy in Patients with COPD. N Eng J Med. 2018;378(18):1671-80.
10. Pavord ID, Lettis S, Anzueto A, Barnes N. Blood eosinophil count and pneumonia risk in patients with chronic obstructive pulmonary disease: a patient-level meta-analysis. Lancet Respir Med. 2016;4(9):731-41.
11. Dicker AJ, Crichton ML, Pumphrey EG, Cassidy AJ, Suarez-Cuartin G, Sibila O, et al. Neutrophil extracellular traps are associated with disease severity and microbiota diversity in patients with chronic obstructive pulmonary disease. J Allergy Clin Immunol. 2018;141(1):117-27.
12. Contoli M, Pauletti A, Rossi MR, Spanevello A, Casolari P, Marcellini A, et al. Long-term effects of inhaled corticosteroids on sputum bacterial and viral loads in COPD. Eur Respir J. 2017;50(4).
13. Wang Z, Bafadhel M, Haldar K, Spivak A, Mayhew D, Miller BE, et al. Lung microbiome dynamics in COPD exacerbations. Eur Respir J. 2016;47(4):1082-92.
14. Faner R, Sibila O, Agusti A, Bernasconi E, Chalmers JD, Huffnagle GB, et al. The microbiome in respiratory medicine: current challenges and future perspectives. Eur Respir J. 2017;49(4).
15. Mayhew D, Devos N, Lambert C, Brown JR, Clarke SC, Kim VL, et al. Longitudinal profiling of the lung microbiome in the AERIS study demonstrates repeatability of bacterial and eosinophilic COPD exacerbations. Thorax. 2018;73(5):422-30.
16. Dicker AJ, Crichton ML, Cassidy AJ, Brady G, Hapca A, Tavendale R, et al. Genetic mannose binding lectin deficiency is associated with airway microbiota diversity and reduced exacerbation frequency in COPD. Thorax. 2018;73(6):510-8.
17. Bafadhel M, McKenna S, Terry S, Mistry V, Reid C, Haldar P, et al. Acute exacerbations of chronic obstructive pulmonary disease: identification of biologic clusters and their biomarkers. Am J Respir Crit Care Med. 2011;184(6):662-71.
18. Wang Z, Singh R, Miller BE, Tal-Singer R, Van Horn S, Tomsho L, et al. Sputum microbiome temporal variability and dysbiosis in chronic obstructive pulmonary disease exacerbations: an analysis of the COPDMAP study. Thorax. 2018;73(4):331-8.

19. Singanayagam A, Schembri S, Chalmers JD. Predictors of mortality in hospitalized adults with acute exacerbation of chronic obstructive pulmonary disease. Ann Am Thorac Soc. 2013;10(2):81-9.
20. Flynn RWV, MacDonald TM, Chalmers JD, Schembri S. The effect of changes to GOLD severity stage on long term morbidity and mortality in COPD. Respir Res. 2018;19(1):249.
21. Soler-Cataluna JJ, Martinez-Garcia MA, Roman Sanchez P, Salcedo E, Navarro M, Ochando R. Severe acute exacerbations and mortality in patients with chronic obstructive pulmonary disease. Thorax. 2005;60(11):925-31.
22. Han MK, Zhou Y, Murray S, Tayob N, Noth I, Lama VN, et al. Lung microbiome and disease progression in idiopathic pulmonary fibrosis: an analysis of the COMET study. Lancet Respir Med. 2014;2(7):548-56.
23. Singh D, Agusti A, Anzueto A, Barnes PJ, Bourbeau J, Celli BR, et al. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Lung Disease: The GOLD Science Committee Report 2019. Eur Respir J. 2019.
24. Fabbri LM, Calverley PM, Izquierdo-Alonso JL, Bundschuh DS, Brose M, Martinez FJ, et al. Roflumilast in moderate-to-severe chronic obstructive pulmonary disease treated with longacting bronchodilators: two randomised clinical trials. Lancet. 2009;374(9691):695-703.
25. Leitao Filho FS, Alotaibi NM, Ngan D, Tam S, Yang J, Hollander Z, et al. Sputum Microbiome is Associated with 1-Year Mortality Following COPD Hospitalizations. Am J Respir Crit Care Med. 2018.
26. Garcia-Vidal C, Almagro P, Romani V, Rodriguez-Carballeira M, Cuchi E, Canales L, et al. Pseudomonas aeruginosa in patients hospitalised for COPD exacerbation: a prospective study. Eur Respir J. 2009;34(5):1072-8.
27. Garcha DS, Thurston SJ, Patel AR, Mackay AJ, Goldring JJ, Donaldson GC, et al. Changes in prevalence and load of airway bacteria using quantitative PCR in stable and exacerbated COPD. Thorax. 2012;67(12):1075-80.
28. Hill AT, Campbell EJ, Hill SL, Bayley DL, Stockley RA. Association between airway bacterial load and markers of airway inflammation in patients with stable chronic bronchitis. Am J Med. 2000;109(4):288-95.
29. Bafadhel M, Haldar K, Barker B, Patel H, Mistry V, Barer MR, et al. Airway bacteria measured by quantitative polymerase chain reaction and culture in patients with stable COPD: relationship with neutrophilic airway inflammation, exacerbation frequency, and lung function. Int J Chron Obstruct Pulmon Dis. 2015;10:1075-83.
30. Gevaert E, Zhang N, Krysko O, Lan F, Holtappels G, De Ruyck N, et al. Extracellular eosinophilic traps in association with Staphylococcus aureus at the site of epithelial barrier defects in patients with severe airway inflammation. J Allergy Clin Immunol. 2017;139(6):1849-60.e6.
31. Ghebre MA, Pang PH, Diver S, Desai D, Bafadhel M, Haldar K, et al. Biological exacerbation clusters demonstrate asthma and chronic obstructive pulmonary disease overlap with distinct mediator and microbiome profiles. J Allergy Clin Immunol. 2018;141(6):2027-36.e12.
32. Hutchison DC. The role of proteases and antiproteases in bronchial secretions. Eur J Respir Dis Suppl. 1987;153:78-85.
33. Rogers GB, Bruce KD, Martin ML, Burr LD, Serisier DJ. The effect of long-term macrolide treatment on respiratory microbiota composition in non-cystic fibrosis bronchiectasis: an analysis from the randomised, double-blind, placebo-controlled BLESS trial. Lancet Respir Med. 2014;2(12):988-96.
34. Alotaibi NM, Chen V, Hollander Z, Leipsic JA, Hague CJ, Murphy DT, et al. Phenotyping and outcomes of hospitalized COPD patients using rapid molecular diagnostics on sputum samples. Int J Chron Obstruct Pulmon Dis. 2019;14:311-9.

Figure Legends

Figure 1. Flow chart of the patients and samples used in this study. Baseline characteristics of the cohort who provided sufficient sputum for microbiome sequencing (n=252) is shown in Table 1

Figure 2. A: Microbiome profiles of all stable samples at the phylum level. **B:** Patient microbiome profiles in COPD at the genus level. The samples are sorted by percentage *Haemophilus* and percentage *Streptococcus* to emphasize the dominant populations.

Figure 3. Relationship between clinical phenotypes and microbiome composition **A:** Alpha diversity measured by Shannon-Wiener Diversity Index and Berger-Parker Dominance Index, is associated with the peripheral blood eosinophil count, p-value from paired T-test, 62 patients were identified as having a peripheral blood eosinophil count of ≤100 cells/μL, 190 had a count of >100 cells/μL. **B:** Average % OTUs at the phyla and genera level for non-eosinophilic patients compared to eosinophilic patients. **C**: % Proteobacteria is negatively correlated with blood eosinophil counts. % Firmicutes is positively correlated with blood eosinophil count. Generalized linear models with binomial errors and a logistic link function were used to show how expected proportions of these phyla are related to blood eosinophil levels.

Figure 4. Relationship between baseline microbiome profiles and long-term survival. **A:** Microbiome subgroups identified at the phylum level, for clarity only samples classed as Firmicutes or Proteobacteria dysbiotic or balanced phyla are shown. **B:** Microbiome subgroups identified at the genera level, only samples classed as *Haemophilus* or *Streptococcus* dysbiotic or balanced genera are shown.

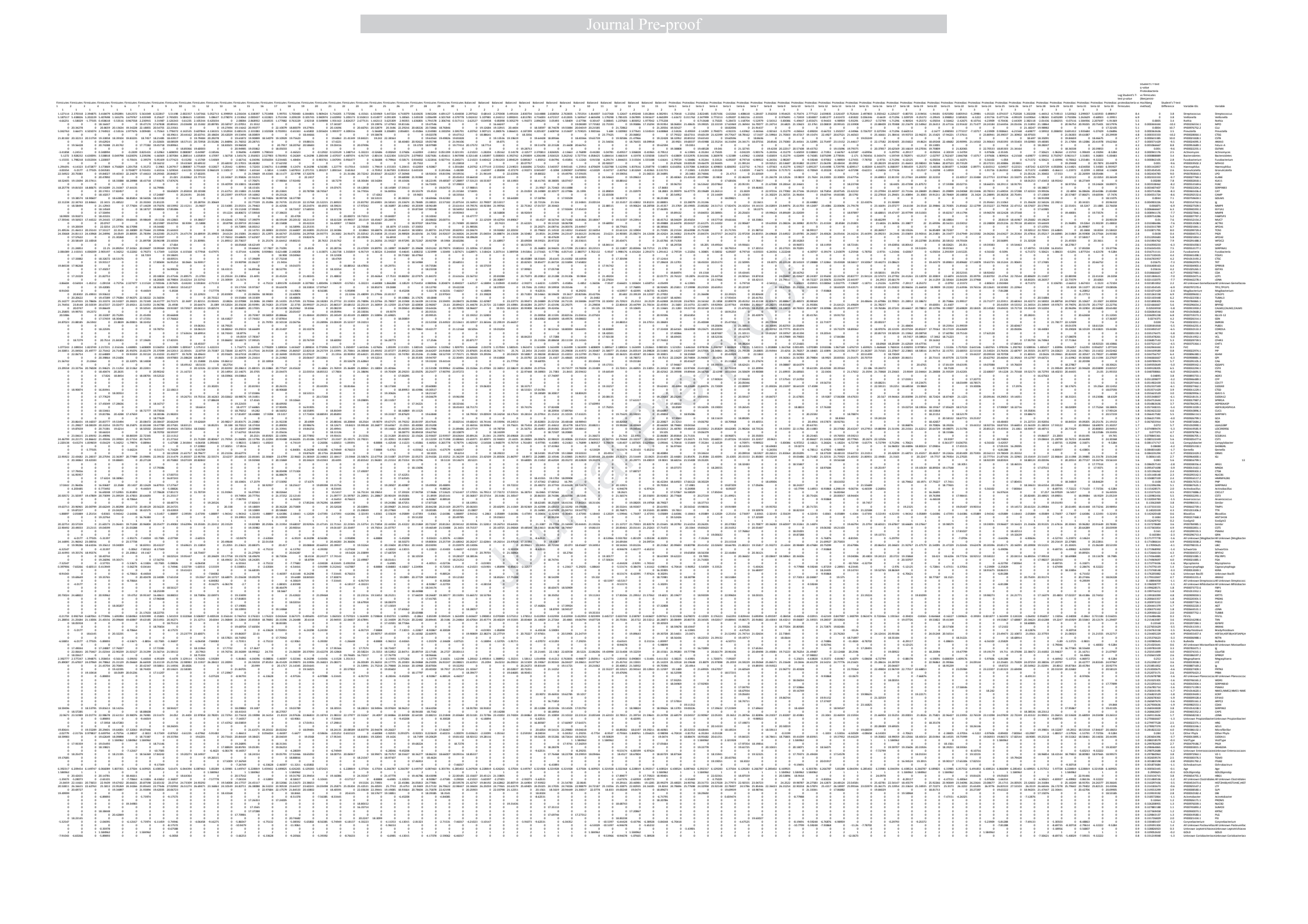
Figure 5. Proteobacteria dominated microbiome profiles are associated with neutrophilic inflammation in patients with COPD. 747 variables (consisting of 613 proteome, 113 microbiome and 21 clinical variables) from n=78 patients were subjected to partial least square discriminant analysis based on three pre-defined microbiome groups. **A:** The scores plot reveals the separation between the three microbiome groups, Firmicutes (n=30, blue dots), balanced (n=15, yellow dots) and Proteobacteria (n=33, red dots), (R^2X=0·17, R^2Y=0·73). **B:** Loadings plot showing which key variables (those closest to the grouping reference points (black dots) drive the separation. Microbiome variables are labelled in red with orange dots, proteins in black with green dots and clinical variables in blue with blue dots. A separate univariate analysis was carried out to confirm the results and to avoid overfitting (Supplementary Table S2).

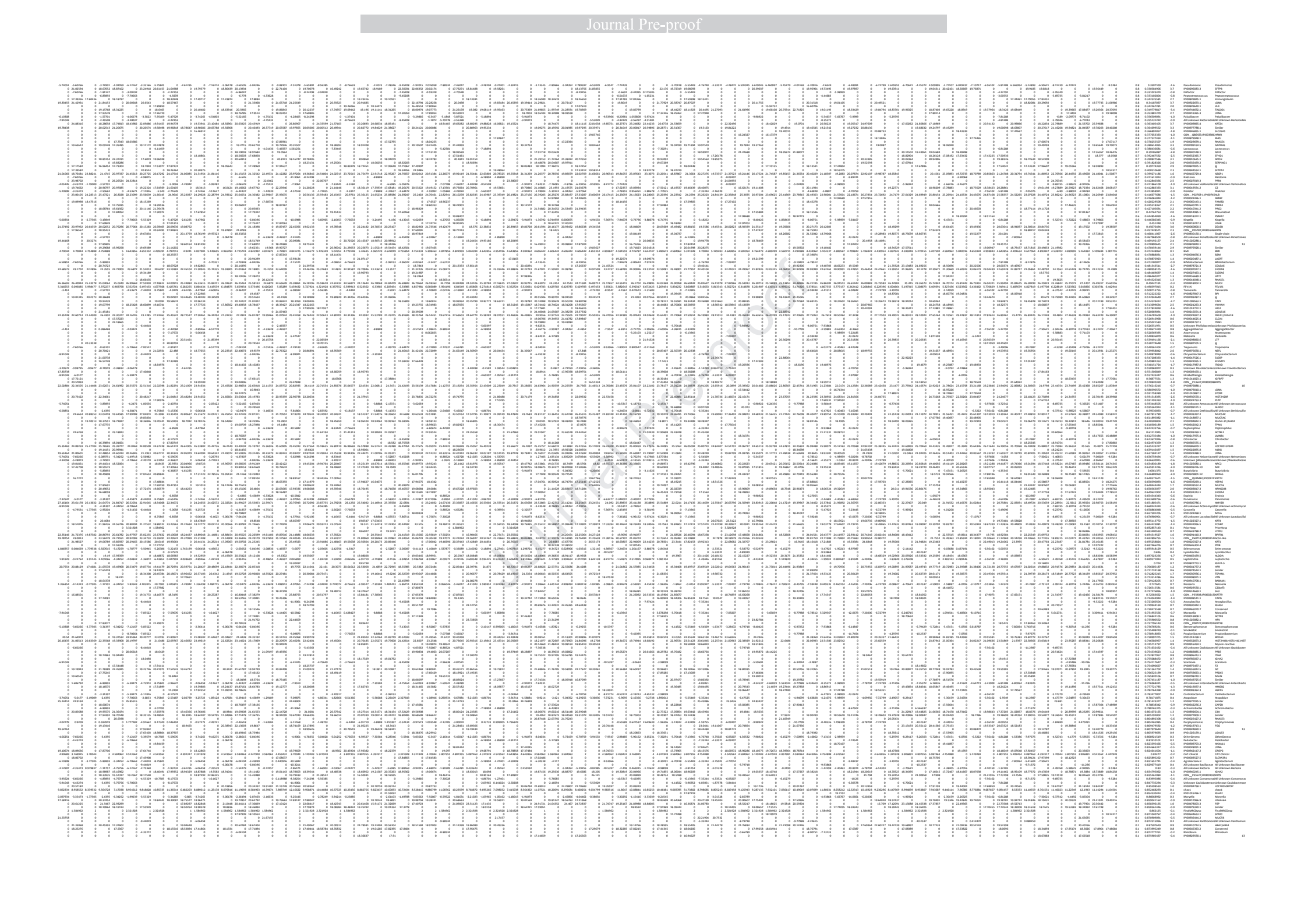
Table 1. Baseline characteristics of the study population (n=252 unique subjects). Abbreviations: ICS= inhaled corticosteroids, BMI= body mass index, MRC= Medical Research Council, FEV_1= forced expiratory volume in 1 second, FVC= forced vital capacity, SGRQ= St Georges Respiratory Questionnaire, CAT= COPD assessment test. *includes medications used in combination with other bronchodilators or inhaled steroids.

Characteristics	**Median (IQR) or n (%)**
N	252
Age	71 (66-78)
Male Gender	153 (60·71%)
BMI	26·47 (24·0-31·02)
ICS use* Long acting beta-agonists* Long acting muscarinic antagonist* Short acting bronchodilators only Oral antibiotics	163 (64·4%) 201 (79·4%) 186 (73·8%) 14 (5·6%) 51 (20.24%)
MRC dyspnoea score	3 (2-4)
Current smokers Ex-smokers Pack years <10 10-20 20-40 40 or more Missing	65 (25·79%) 184 (73·02%) 12 (4·76%) 37 (14·68%) 94 (37·30%) 107 (42·46%) 2 (0·79%)
Exacerbation frequency (year prior to the study) 0 1 2 3 or more Severe exacerbation requiring hospitalization (year prior to the study)	 52 (20·63%) 48 (19·0%) 40 (15·8%) 112 (44·3%) 57 (22·62%)
Daily Sputum volume (mL)	10·0 (4·75-20·0)
Spirometry FEV_1 (L) FEV_1 (% predicted) FVC FEV_1/FVC	 1·50 (1·06-1·95) 65·35 (49·75-80·00) 2·87 (2·11-3·61) 53·00 (44·75-61·02)
GOLD 2017 A B C	 58 (23·02%) 32 (12·6%) 52 (20·6%)

D	110 (43·5%)
SGRQ total score CAT total score	47·82 (31·93-64·58) 19 (14-25)

507




Journal Pre-proof

e

tacrae]

acrae
tacrae

ddf [acr4

rae

rcae

]

tacrae

ae

i
TJH6],HIST2H4D],HIST1H4A],HIST2H4A],HIST1H4I],HIST1H4D],HIST2H4E],HIST2H4L],HIST1H4F],HIST1H4H],HIST4H4],HIST2H4

ycae
ae
rcae

es

eae

s

:eae

eae

eae
vibrionaceae

ae
e]

taceae

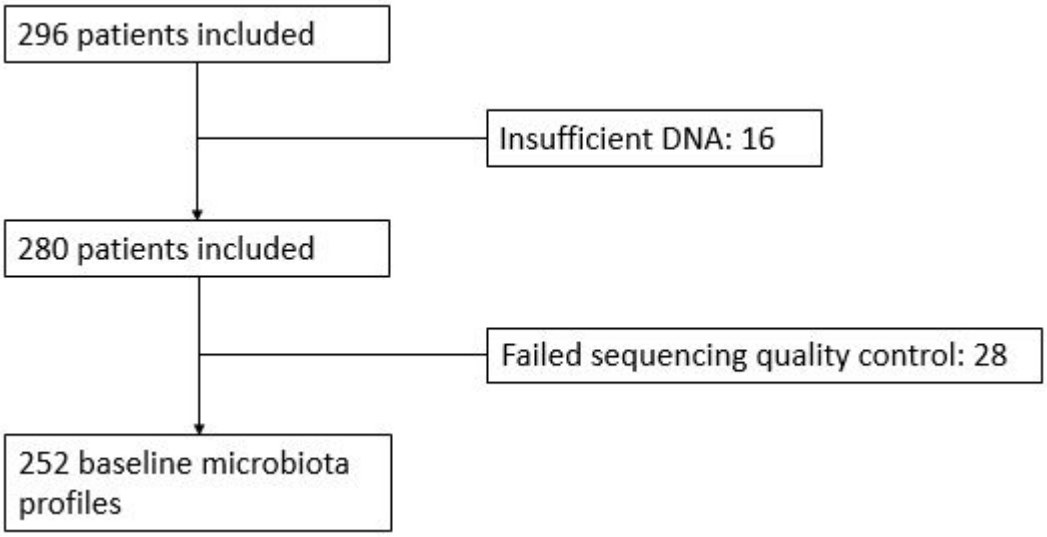
les

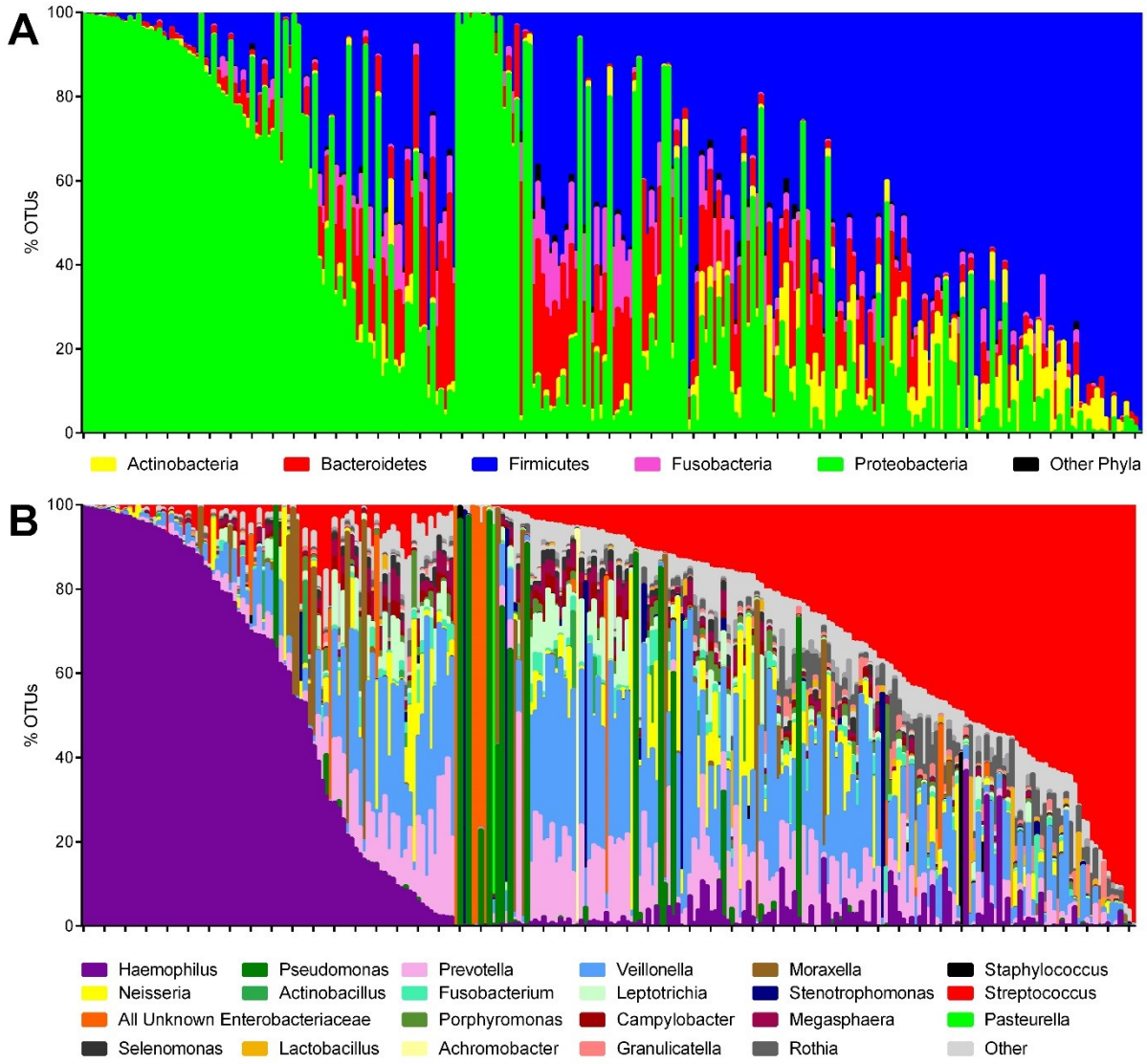
T1H4J;HIST1H4D;HIST1H4A;HIST2H4A;HIST1H4I;HIST1H4K;HIST1H4E;HIST1H4L;HIST1H4F;HIST1H4H;HIST4H4;HIST1H4
raceae

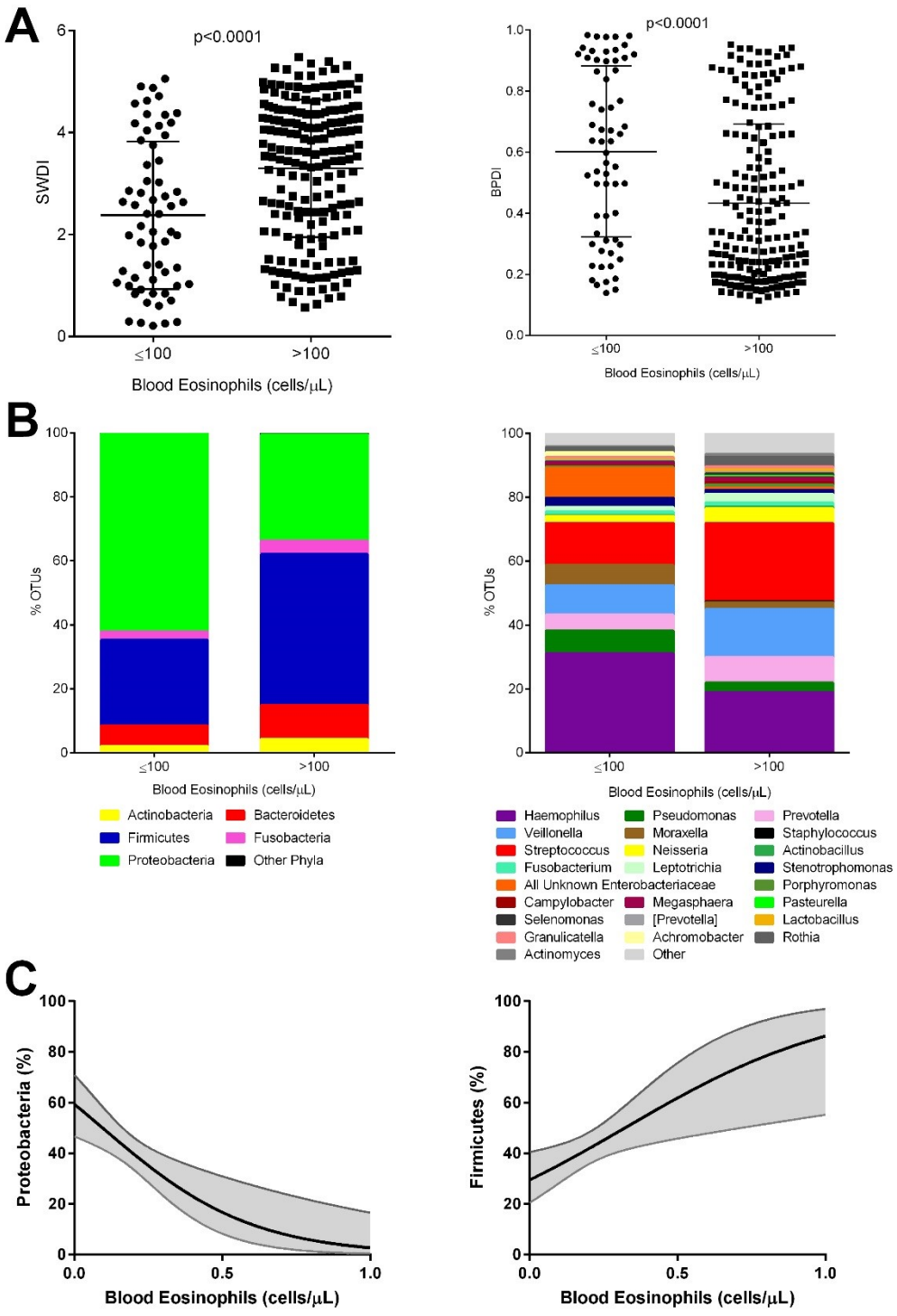
eriaceae

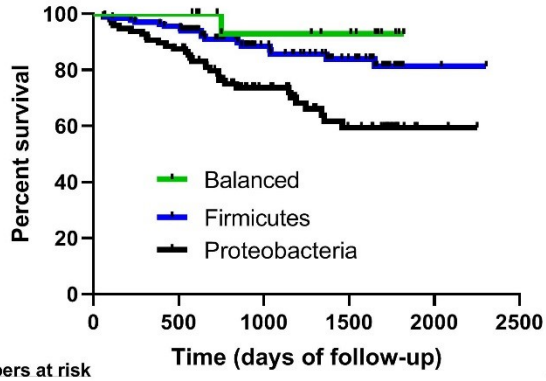
daceae
eae

adaceae

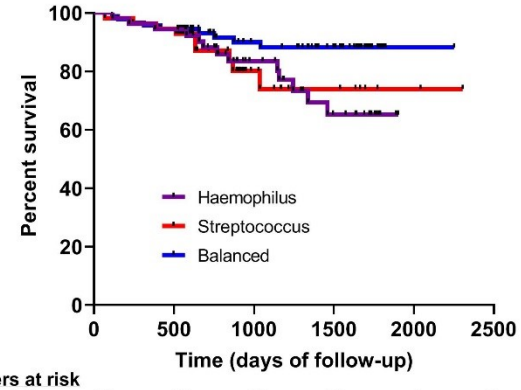



A
100
80
60
40
20
0
% OTUs
Actinobacteria
Bacteroidetes
Firmicutes
Fusobacteria
Proteobacteria
Other Phyla
B
100
80
60
40
20
0
% OTUs
Haemophilus
Neisseria
All Unknown Enterobacteriaceae
Selenomonas
Pseudomonas
Actinobacillus
Lactobacillus
Prevotella
Fusobacterium
Porphyromonas
Achromobacter
Veillonella
Leptotrichia
Campylobacter
Granulicatella
Moraxella
Stenotrophomonas
Megasphaera
Rothia
Staphylococcus
Streptococcus
Pasteurella
Other


A
p<0.0001
SWDI
Blood Eosinophils (cells/μL)
p<0.0001
BPDI
Blood Eosinophils (cells/μL)
B
% OTUs
Blood Eosinophils (cells/μL)
Actinobacteria
Bacteroidetes
Firmicutes
Fusobacteria
Proteobacteria
Other Phyla
% OTUs
Blood Eosinophils (cells/μL)
Haemophilus
Pseudomonas
Prevotella
Veillonella
Moraxella
Staphylococcus
Streptococcus
Neisseria
Actinobacillus
Fusobacterium
Leptotrichia
Stenotrophomonas
All Unknown Enterobacteriaceae
Porphyromonas
Campylobacter
Megasphaera
Pasteurella
Selenomonas
[Prevotella]
Lactobacillus
Granulicatella
Achromobacter
Rothia
Actinomyces
Other
C
Proteobacteria (%)
Blood Eosinophils (cells/μL)
Firmicutes (%)
Blood Eosinophils (cells/μL)



Numbers at risk						
Firmicutes	135	130	64	44	3	0
Balanced	18	18	13	11	0	0
Proteobacteria	96	85	46	27	4	0



Numbers at risk						
Haemophilus	53	51	29	16	2	0
Streptococcus	56	54	15	8	3	0
Balanced	91	87	52	38	0	0